14th International Conference on Information Fusion
Chicago, Illinois, USA, July 5-8, 2011

Nonlinear filtering update phase via the Single Point Truncated Unscented Kalman Filter

Ángel F. García-Fernández*, Mark R. Morelande†, Jesús Grajal*
*Dpto. Señales, Sistemas y Radiocomunicaciones, Universidad Politécnica de Madrid, Spain
†Melbourne Systems Laboratory, The University of Melbourne, Australia
Emails: agarcia@gmr.ssr.upm.es, mrmore@unimelb.edu.au, jesus@gmr.ssr.upm.es

***Abstract*—A fast algorithm to approximate the first two moments of the posterior probability density function (pdf) in nonlinear non-Gaussian Bayesian filtering is proposed. If the pdf of the measurement noise has a bounded support and the measurement function is continuous and bijective, we can use a modified prior pdf that meets Bayes' rule exactly. The central idea of this paper is that a Kalman filter applied to a modified prior distribution can improve the estimate given by the conventional Kalman filter. In practice, bounded support is not required and the modification of the prior is accounted for by adding an extra-point to the set of sigma-points used by the unscented Kalman filter.**

***Index Terms*—Bayes' rule, Kalman filter, nonlinear filtering**

I. Introduction

A Bayesian estimator of a dynamic system consists of two phases that are performed at every time step: prediction and update [1]. In the prediction phase, the filter approximates the pdf of the current state given all previous measurements. This pdf is referred to as the prior pdf. In the update stage, Bayes' rule is applied to the prior pdf using the measurement equation yielding the posterior pdf. The minimum mean square error (MMSE) estimator involves the knowledge of the posterior pdf. In linear Gaussian models, the posterior pdf is calculated analytically by the Kalman filter (KF). In non-linear non-Gaussian models there is no closed-form solution to the posterior pdf and it must be approximated.

In this paper, we focus on the update phase, i.e., we develop an algorithm that approximates the posterior assuming the prior is known. A common family of algorithms used to approximate the posterior pdf are particle filters (PFs) [2]. PFs provide arbitrarily accurate approximation given sufficient computational resources but can perform poorly if computations are limited. A suitable alternative in cases where the posterior can be accurately represented by its first two moments is to use a Gaussian approximation. One way of computing such an approximation is to use the KF. When the measurement equation is linear, the prior and the pdf of measurement noise are Gaussian, the Gaussian pdf represented by the mean and covariance matrix given by KF update equation is the result of applying the Bayes' rule [3]. KF can also be applied when the measurement model is nonlinear. Yet, in this case, KF does not perform directly the Bayes' rule in the update phase and then, convergence to the true posterior pdf does not occur. This is its main difference with PFs as they converge to the true posterior if the number of particles tends to infinity. Nevertheless, if nonlinear effects are mild, KF is expected to give an accurate enough representation of the true posterior with low computational burden.

Another problem with Kalman filtering in nonlinear problems is that, in general, the moments required in the KF recursion cannot be computed exactly. Approximations such as the extended Kalman filter (EKF) [1], unscented Kalman filter (UKF) [4] or cubature Kalman filter (CKF) [5] have been proposed. These conventional Kalman-filter-type approaches to nonlinear filtering are different ways of computing the integrals of the required moments. However, it is known that measurement nonlinearities become significant when the measurement noise variance is small compared to the prior variance [6]. Then, in such a condition, all these algorithms are expected to approximate the posterior poorly.

Instead, the approach proposed in this paper departs from the previous techniques and aims to improve on KF performance when nonlinearities are significant. It consists of approximating the pdf of the measurement noise by one with bounded support. This implies that the likelihood has a bounded support too if the measurement function is continuous and bijective. Thus, the exact Bayes' rule can also be obtained by using a modified prior, which is a mixture of the original prior and a truncated version of it. The motivation behind this is that the variance of the modified prior is expected to be lower than the variance of the original prior so that the Kalman filter applied to this modified prior should work better in principle. The way we propose to put this idea into practice is called single point truncated unscented Kalman filter (SPTUKF). It uses the same set of sigma-points as the UKF plus one extra point to approximate the moments required to apply the KF to the modified prior distribution.

The remainder of this paper is organised as follows. In Section II, we review the update stage of the KF for nonlinear measurement models. In Section III, we present the truncated Kalman filter (TKF). In Section IV, we present the single point truncated unscented Kalman filter as an approximation of the TKF. Simulation examples analysing the update phase of the filter and tracking performance are provided in Section V. Finally, conclusions are drawn in Section VI.

II. Kalman filter for nonlinear measurement models

In this section, the update phase of a Kalman filter for nonlinear measurement models is analysed. For ease of explanation,

we do not include the time index and the conditioning on past measurements. Therefore, the prior pdf (the pdf of the current state given all previous measurements) is denoted by $p_0(\cdot)$. We assume the following measurement equation:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\eta} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the state, $\mathbf{z} \in \mathbb{R}^{n_z}$ is the measurement, $\mathbf{h}(\cdot)$ is a nonlinear function and $\boldsymbol{\eta}$ is the zero mean measurement noise with covariance matrix $\mathbf{R}$. . From the Bayesian point of view, the minimum mean square error estimator $\hat{\mathbf{x}}_{pos}$ of $\mathbf{x}$ given $\mathbf{z}$ is [1]:

$$\hat{\mathbf{x}}_{pos} = \mathsf{E}[\mathbf{x}|\mathbf{z}] \tag{2}$$

whose mean square error matrix is

$$\mathbf{P}_{pos} = \mathsf{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_{pos})(\mathbf{x} - \hat{\mathbf{x}}_{pos})^T |\mathbf{z}\right] \tag{3}$$

which is measurement dependent and its trace is the minimum mean square error we can achieve as $\hat{\mathbf{x}}_{pos}$ is the MMSE estimator. Calculating (2) and (3) is our ultimate objective as they are the first two moments of the posterior. However, they are very difficult to calculate as we need to know the posterior.

Conversely, the updated mean of the KF provides the linear MMSE estimator of $\mathbf{x}$ given $\mathbf{z}$, $\hat{\mathbf{x}}_{u,0}$ [3]:

$$\hat{\mathbf{x}}_{u,0} = \hat{\mathbf{x}}_{p,0} + \boldsymbol{\Psi}_0 \mathbf{S}_0^{-1}(\mathbf{z} - \hat{\mathbf{z}}_0) \tag{4}$$

where $\hat{\mathbf{x}}_{p,0}$ is the mean of the prior distribution of $\mathbf{x}$ and

$$\hat{\mathbf{z}}_0 = \mathsf{E}[\mathbf{z}] = \int \mathsf{E}[\mathbf{z}|\mathbf{x}]\, p_0(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{5}$$

$$\mathbf{S}_0 = \mathsf{cov}[\mathbf{z}] = \int \mathsf{E}\left[(\mathbf{z} - \hat{\mathbf{z}}_0)(\mathbf{z} - \hat{\mathbf{z}}_0)^T |\mathbf{x}\right] p_0(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{6}$$

$$\boldsymbol{\Psi}_0 = \mathsf{cov}[\mathbf{x}, \mathbf{z}] = \int \mathsf{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_{p,0})(\mathbf{z} - \hat{\mathbf{z}}_0)^T |\mathbf{x}\right] p_0(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{7}$$

The updated covariance matrix given by the KF, $\mathbf{P}_{u,0}$, is

$$\mathbf{P}_{u,0} = \mathbf{P}_{p,0} - \boldsymbol{\Psi}_0 \mathbf{S}_0^{-1} \boldsymbol{\Psi}_0^T \tag{8}$$

where $\mathbf{P}_{p,0}$ is the covariance of the prior distribution of $\mathbf{x}$. As shown in [3], $\mathbf{P}_{u,0}$ is the mean square error matrix averaged over all possible measurements even in a nonlinear set-up:

$$\mathbf{P}_{u,0} = \mathsf{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_{u,0})(\mathbf{x} - \hat{\mathbf{x}}_{u,0})^T\right] \tag{9}$$

In filtering problems, $\hat{\mathbf{x}}_{u,0}$ and $\mathbf{P}_{u,0}$ are usually regarded as approximations of the first two moments of the posterior, $\hat{\mathbf{x}}_{pos}$ and $\mathbf{P}_{pos}$. The reason for this is that, when the prior $p_0(\cdot)$ and the measurement noise are Gaussian, and $\mathbf{h}(\cdot)$ is a linear function, $\hat{\mathbf{x}}_{pos} = \hat{\mathbf{x}}_{u,0}$ and $\mathbf{P}_{pos} = \mathbf{P}_{u,0}$ having a closed form solution [3]. Also, if the measurement function is nonlinear but one assumes that the variable $(\mathbf{x}, \mathbf{z})$ is jointly Gaussian, KF update equations are equivalent to the Bayes' rule [7]. However, in this case, there is no guarantee that $\hat{\mathbf{x}}_{u,0}$ and $\mathbf{P}_{u,0}$ are close to $\hat{\mathbf{x}}_{pos}$ and $\mathbf{P}_{pos}$. In fact, it is known that measurement nonlinearities become significant when the measurement noise variance is small compared to the prior variance [6]. Then, in such a condition, all the algorithms that are based on Kalman filtering, such as the UKF, CKF or EKF are expected to approximate the posterior poorly.

Once the shortcomings of Kalman filtering for nonlinear measurement models have been discussed, in the rest of the paper, we will develop an algorithm that improves the approximation of the first two moments of the posterior given by conventional KF-type algorithms.

III. Truncated Kalman Filter

We assume that $\mathbf{h}(\cdot)$ is a continuous, bijective function of some elements of the state. That is, we can write $\mathbf{x} = \left[\mathbf{a}^T, \mathbf{b}^T\right]^T$, where $\mathbf{a} \in \mathbb{R}^{n_a}$, $\mathbf{b} \in \mathbb{R}^{n_b}$ and $n_x = n_a + n_b = n_z + n_b$, such that:

$$\mathbf{z} = \mathbf{h}(\mathbf{a}) + \boldsymbol{\eta} \tag{10}$$

We also assume that the pdf of $\boldsymbol{\eta}$ has a bounded, connected support:

$$p_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = 0 \quad \text{if} \quad \boldsymbol{\eta} \notin I_{\boldsymbol{\eta}} \subset \mathbb{R}^{n_z} \tag{11}$$

where $I_{\boldsymbol{\eta}}$ is an n_z-dimensional connected region. We want to clarify that the condition that the measurement noise has a bounded support has to be met to develop and motivate the theory of the TKF. Nevertheless, in a practical implementation, this condition does not have to be met.

As the pdf of $\boldsymbol{\eta}$ has a bounded, connected support, the pdf of the measurement conditioned on the state can be written as

$$\begin{aligned} p(\mathbf{z}|\mathbf{x}) &= p(\mathbf{z}|\mathbf{a}) \\ &= p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a})) \\ &= p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\, \chi_{I_{\boldsymbol{\eta}}}(\mathbf{z} - \mathbf{h}(\mathbf{a})) \end{aligned} \tag{12}$$

where $\chi_{I_{\boldsymbol{\eta}}}(\cdot)$ is the indicator function on the subset $I_{\boldsymbol{\eta}}$.

As $\mathbf{h}(\cdot)$ is continuous and bijective, (12) can be written as:

$$p(\mathbf{z}|\mathbf{x}) = p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\, \chi_{I_{\mathbf{x}}(\mathbf{z})}(\mathbf{x}) \tag{13}$$

where

$$\begin{aligned} I_{\mathbf{x}}(\mathbf{z}) =& I_{\mathbf{a}}(\mathbf{z}) \times \mathbb{R}^{n_b} \\ =& \left\{ \mathbf{x} \,\middle|\, \mathbf{x} = \left[\left(\mathbf{h}^{-1}(\mathbf{z} - \boldsymbol{\eta})\right)^T, \mathbf{b}^T\right]^T, \boldsymbol{\eta} \in I_{\boldsymbol{\eta}}, \mathbf{b} \in \mathbb{R}^{n_b} \right\} \end{aligned} \tag{14}$$

The sets $I_{\mathbf{x}}(\mathbf{z})$ and $I_{\mathbf{a}}(\mathbf{z})$ depend on the current measurement $\mathbf{z}$, the support of the measurement noise $I_{\boldsymbol{\eta}}$ and the inverse function of $\mathbf{h}(\cdot)$. The posterior pdf of $\mathbf{x}$ applying Bayes' rule and (13) is:

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x}) \cdot p_0(\mathbf{x}) \tag{15}$$

$$= p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\, \chi_{I_{\mathbf{x}}(\mathbf{z})}(\mathbf{x})\, p_0(\mathbf{x}) \tag{16}$$

A usual approach to approximate the first two moments of the posterior is to apply a Kalman filter using $p_0(\cdot)$ as the prior (KF). However, (16) can be rewritten as

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})\, p_2(\mathbf{x}; \mathbf{z}) \tag{17}$$

where

$$p_2(\mathbf{x}; \mathbf{z}) = \alpha_{sp} p_1(\mathbf{x}; \mathbf{z}) + (1 - \alpha_{sp})\, p_0(\mathbf{x}) \tag{18}$$

where $\alpha_{sp} \in [0, 1]$ and

$$p_1(\mathbf{x}; \mathbf{z}) = \frac{1}{\varepsilon_1} p_0(\mathbf{x})\, \chi_{I_{\mathbf{x}}(\mathbf{z})}(\mathbf{x}) \tag{19}$$

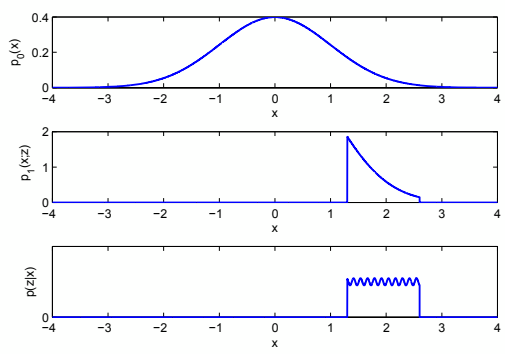

Figure 1. Graphical representation of $p_0(\mathbf{x})$, $p_1(\mathbf{x};\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$ in one dimension. If the likelihood has a bounded support, the Bayes' rule is exactly met by the "prior" pdfs: $p_0(\mathbf{x})$, $p_1(\mathbf{x};\mathbf{z})$ or any mixture of both $p_2(\mathbf{x};\mathbf{z})$.

and ε_1 is a normalising constant and $p_1(\cdot)$ is a modified "prior" pdf, which is parameterised by $\mathbf{z}$. We should note that $p_1(\cdot)$ is a truncated version of $p_0(\cdot)$ and the result of applying Bayes' rule, equations (16) and (17), is the same if we use $p_0(\cdot)$, $p_1(\cdot)$ or any mixture of both $p_2(\cdot)$ as the prior. This is proven in Appendix A and illustrated in Fig. 1.

Equation (17) suggests an alternative approach to KF using the mixture $p_2(\cdot)$ as the prior, the truncated KF (TKF). As indicated in Section II, it is known that measurement nonlinearities become significant when the measurement noise variance is small compared to the prior variance [6]. Then, the use of $p_2(\cdot)$ instead of $p_0(\cdot)$ aims to reduce the variance of the prior pdf, so that the effect of nonlinearities decreases, while at the same time keeping the output of Bayes' rule unaltered. It is expected this will enable the TKF to achieve better performance than the KF in nonlinear set-ups. The degree of freedom given by α_{sp} is of paramount importance in the practical implementation of the algorithm.

IV. Single Point Truncated Unscented Kalman Filter

As we have indicated in the previous section, the use of the TKF could provide performance benefits over the KF. However, the measurement noise is not usually bounded so the algorithm explained before cannot be directly applied to practical problems in nonlinear filtering. Nonetheless, in the same way that KF and approximations are used in nonlinear/non-Gaussian systems when the equations KF solves are only optimal for linear and Gaussian systems, we can approximate the TKF when the measurement noise pdf is not bounded. As we will demonstrate in practical scenarios in Section V, this approximation is usually better than the one given by KF and approximations when nonlinear effects become important.

Then, in this section we address the two approximations we have to make to apply the ideas of the TKF to practical problems: the first one to approximate the first two moments of $p_2(\cdot)$ and the second one to approximate the moments (5), (6) and (7) with respect to the modified prior. In addition, we also explain how to calculate α_{sp} such that the algorithm works properly. It will be shown, that α_{sp} will be high when the approximation of $p_1(\cdot)$ is good and it will be low when the approximation of $p_1(\cdot)$ is inaccurate.

A. Approximating the first two moments of $p_2(\cdot)$

In this section, we approximate the first two moments of $p_2(\mathbf{x};\mathbf{z}) = p_2(\mathbf{a},\mathbf{b};\mathbf{z})$. Firstly, we should realise that due to the partitioning of the state, the first two moments of $p_0(\mathbf{a},\mathbf{b})$ and $p_2(\mathbf{a},\mathbf{b};\mathbf{z})$ can be written as:

$$\hat{\mathbf{x}}_{p,j} = \begin{bmatrix} \mathsf{E}[\mathbf{a}] \\ \mathsf{E}[\mathbf{b}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{a,j} \\ \boldsymbol{\mu}_{b,j} \end{bmatrix} \tag{20}$$

$$\mathbf{P}_{p,j} = \begin{bmatrix} \mathsf{cov}[\mathbf{a}] & \mathsf{cov}[\mathbf{a},\mathbf{b}] \\ \mathsf{cov}[\mathbf{b},\mathbf{a}] & \mathsf{cov}[\mathbf{b}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{a,j} & \boldsymbol{\Sigma}_{ab,j} \\ \boldsymbol{\Sigma}_{ab,j}^T & \boldsymbol{\Sigma}_{b,j} \end{bmatrix} \tag{21}$$

where $\hat{\mathbf{x}}_{p,j}$ and $\mathbf{P}_{p,j}$ are the mean and covariance matrix of $p_j(\cdot)$ for $j \in \{0,2\}$, respectively. The truncation is directly applied only to $\mathbf{a}$ and not $\mathbf{b}$. Then, considering that

$$p_0(\mathbf{x}) = p_0(\mathbf{a},\mathbf{b}) = p_0(\mathbf{b}|\mathbf{a})\, p_0(\mathbf{a}) \tag{22}$$

the truncated pdf $p_1(\mathbf{x};\mathbf{z})$ can be written as

$$p_1(\mathbf{x};\mathbf{z}) = p_0(\mathbf{b}|\mathbf{a})\, p_1(\mathbf{a};\mathbf{z}) \tag{23}$$

where we highlight that the conditional pdf of $\mathbf{b}$ given $\mathbf{a}$ is the same for $p_1(\mathbf{x};\mathbf{z})$ and $p_0(\mathbf{x})$ because the truncation is applied to $\mathbf{a}$. Then, (18) becomes:

$$p_2(\mathbf{x};\mathbf{z}) = p_0(\mathbf{b}|\mathbf{a})\, p_2(\mathbf{a};\mathbf{z}) \tag{24}$$

where

$$p_2(\mathbf{a};\mathbf{z}) = \alpha_{sp} p_1(\mathbf{a};\mathbf{z}) + (1-\alpha_{sp})\, p_0(\mathbf{a}) \tag{25}$$

The SP-TUKF uses the following approximation of $p_2(\mathbf{a};\mathbf{z})$:

$$p_2(\mathbf{a};\mathbf{z}) \approx \alpha_{sp}\delta(\mathbf{a} - \tilde{\mathbf{a}}(\mathbf{z})) + (1-\alpha_{sp})\, \tilde{p}_0(\mathbf{a}) \tag{26}$$

where $\delta(\cdot)$ is the Dirac delta, $\tilde{\mathbf{a}}(\mathbf{z})$ is a sigma-point that represents $p_1(\cdot)$, whose selection is addressed in Section IV-C, and $\tilde{p}_0(\cdot)$ is the unscented approximation of $p_0(\cdot)$ [8], [9]:

$$p_0(\mathbf{a}) \approx \tilde{p}_0(\mathbf{a}) = \sum_{i=1}^{N_s} w^i \delta(\mathbf{a} - \mathbf{a}^i) \tag{27}$$

where N_s is the number of sigma-points and $\mathbf{a}^i$ is the ith sigma-point and w^i is the weight of sigma-point i. The selection of the sigma-points and their weights can be found in [4]. In our implementations, we use $N_s = 2n_x + 1$ sigma-points and the weight of sigma-point located on the mean of $p_0(\cdot)$ is $1/3$ as suggested in Section IV of [4]. The reason why a single sigma-point is used to represent $p_1(\cdot)$ as well as the proper selection of α_{sp} are addressed in Section IV-C.

Then, if we call

$$\left(\mathcal{A}_2^1, \ldots, \mathcal{A}_2^{N_s+1}\right) = \left(\mathbf{a}^1, \ldots, \mathbf{a}^{N_s}, \tilde{\mathbf{a}}(\mathbf{z})\right) \tag{28}$$

and

$$\left(w_2^1, ..., w_2^{N_s+1}\right) = \left((1-\alpha_{sp})\, w^1, ..., (1-\alpha_{sp})\, w^{N_s}, \alpha_{sp}\right) \tag{29}$$

The SP-TUKF approximation of $p_2(\mathbf{a};\mathbf{z})$ can be written as:

$$p_2(\mathbf{a};\mathbf{z}) = \sum_{i=1}^{N_s+1} w_2^i \delta\left(\mathbf{a} - \mathcal{A}_2^i\right) \tag{30}$$

The approximated mean and covariance matrix of $p_2(\mathbf{a};\mathbf{z})$ are:

$$\boldsymbol{\mu}_{a,2} = \sum_{i=1}^{N_s+1} w_2^i \mathcal{A}_2^i \tag{31}$$

$$\boldsymbol{\Sigma}_{a,2} = \sum_{i=1}^{N_s+1} w_2^i \left(\mathcal{A}_2^i - \boldsymbol{\mu}_{a,2}\right)\left(\mathcal{A}_2^i - \boldsymbol{\mu}_{a,2}\right)^T \tag{32}$$

Based on (31) and (32), the rest of the elements of the first two moments of $p_2(\mathbf{a},\mathbf{b};\mathbf{z})$, given by (20) and (21), can be found to be

$$\boldsymbol{\mu}_{b,2} = \boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1}\left(\boldsymbol{\mu}_{a,2} - \boldsymbol{\mu}_{a,0}\right) \tag{33}$$

$$\boldsymbol{\Sigma}_{ab,2} = \boldsymbol{\Sigma}_{a,2}\left(\boldsymbol{\Sigma}_{a,0}^{-1}\right)^T \boldsymbol{\Sigma}_{ab,0} \tag{34}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{b,2} =& \boldsymbol{\Gamma} - \left(\boldsymbol{\mu}_{b,2} - \boldsymbol{\mu}_{b,0}\right)\left(\boldsymbol{\mu}_{b,2} - \boldsymbol{\mu}_{b,0}\right)^T + \\ & \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1}\left[\boldsymbol{\Sigma}_{a,2} + \left(\boldsymbol{\mu}_{a,2} - \boldsymbol{\mu}_{a,0}\right)\left(\boldsymbol{\mu}_{a,2} - \boldsymbol{\mu}_{a,0}\right)^T\right] \cdot \\ & \left(\boldsymbol{\Sigma}_{a,0}^{-1}\right)^T \boldsymbol{\Sigma}_{ab,0}\end{aligned} \tag{35}$$

where

$$\boldsymbol{\Gamma} \triangleq \mathsf{cov}\left[\mathbf{b}\,|\mathbf{a}\right] = \boldsymbol{\Sigma}_{b,0} - \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1} \boldsymbol{\Sigma}_{ab,0} \tag{36}$$

In Appendix B, we show the procedure to calculate $\boldsymbol{\mu}_{b,2}$. The rest of the moments are calculated in a similar way.

B. Approximation of the prior moments

In this subsection, we describe how we approximate integrals (5), (6) and (7) for $p_2(\cdot)$. We will use the unscented transformation (UT) for conditionally linear models [10].

Substituting (24) into (5)

$$\hat{\mathbf{z}}_2 = \int \mathbf{h}(\mathbf{a})\, p_0(\mathbf{b}\,|\mathbf{a})\, p_2(\mathbf{a})\, \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \tag{37}$$

Integrating out $\mathbf{b}$ in (37):

$$\hat{\mathbf{z}}_2 = \int \mathbf{h}(\mathbf{a})\, p_2(\mathbf{a})\, \mathrm{d}\mathbf{a} \tag{38}$$

Using (28) and (29), the transformed sigma-points are calculated as

$$\mathcal{Z}_2^i = \mathbf{h}\left(\mathcal{A}_2^i\right), \quad i = 1, ..., N_s + 1 \tag{39}$$

and the UT approximation to (38) is

$$\hat{\mathbf{z}}_2 = \sum_{i=1}^{N_s+1} w_2^i \mathcal{Z}_2^i \tag{40}$$

The UT approximation to (6) is calculated using (39) [10]:

$$\mathbf{S}_2 = \mathbf{R} + \sum_{i=1}^{N_s+1} w_2^i \left(\mathcal{Z}_2^i - \hat{\mathbf{z}}_2\right)\left(\mathcal{Z}_2^i - \hat{\mathbf{z}}_2\right)^T \tag{41}$$

where $\mathbf{R}$ is the covariance matrix of the measurement noise.

Let $\mathcal{X}_2^i = \left[\left(\mathcal{A}_2^i\right)^T, \left(\boldsymbol{\upsilon}_b\left(\mathcal{A}_2^i\right)\right)^T\right]^T$ for $i = 1, ..., N_s + 1$ where $\boldsymbol{\upsilon}_b(\cdot)$ is given in Appendix B. Then, (7) can be approximated as [10]

$$\boldsymbol{\Psi}_2 = \sum_{i=1}^{N_s+1} w_2^i \left(\mathcal{X}_2^i - \hat{\mathbf{x}}_{p,2}\right)\left(\mathcal{Z}_2^i - \hat{\mathbf{z}}_2\right)^T \tag{42}$$

C. Selection of the parameters $\tilde{\mathbf{a}}(\mathbf{z})$ and α_{sp}

As we indicated in Section II, conventional KF-type algorithms perform badly when the variance of the prior is large compared with the variance of the measurement noise because nonlinearities become significant in this situation [6]. Then, the TKF improvement is expected to be meaningful when the size of the region $I_{\mathbf{a}}(\mathbf{z})$ is small. In such a case, the approximation of the truncated pdf $p_1(\mathbf{a};\mathbf{z})$ by a single sigma-point $\tilde{\mathbf{a}}(\mathbf{z})$ is accurate enough to improve the performance of the filter with nonlinear measurement models as we demonstrate with numerical examples in Section V.

Then, we can use the degree of freedom given by α_{sp} in the following way. As the variance of $p_1(\cdot)$ decreases, the accuracy of the KF applied to $p_1(\cdot)$ increases over the accuracy of the KF applied to $p_0(\cdot)$. Also, the approximation of $p_1(\mathbf{a};\mathbf{z})$ by a single sigma-point $\tilde{\mathbf{a}}(\mathbf{z})$ improves when the size of $I_{\mathbf{a}}(\mathbf{z})$ gets smaller, which implies that the variance of $p_1(\cdot)$ decreases. Thus, due to these two reasons, α_{sp}, which indicates the weight of $p_1(\cdot)$ in the mixture $p_2(\cdot)$, should be chosen such that it favours $p_1(\cdot)$ when its variance is low compared to the variance of $p_0(\cdot)$. Therefore, firstly, we provide an approximation for the variance of $p_1(\cdot)$ and, then, we give a heuristic equation for α_{sp} that meets these requirements.

Using a first-order Taylor series of $\mathbf{h}(\cdot)$ around $\tilde{\mathbf{a}}(\mathbf{z}) = \arg\max_{\mathbf{a}} p(\mathbf{z}\,|\mathbf{a})$, i.e. the value that maximises the likelihood, we can approximate the region

$$I_{\mathbf{a}}(\mathbf{z}) \approx \left\{\mathbf{a}\,\middle|\,\mathbf{a} = \tilde{\mathbf{a}}(\mathbf{z}) - \tilde{\mathbf{H}}^{-1}\boldsymbol{\eta},\, \boldsymbol{\eta} \in I_{\boldsymbol{\eta}}\right\} \tag{43}$$

where

$$\tilde{\mathbf{H}} = \left[\nabla_{\mathbf{a}}\mathbf{h}^T\left(\tilde{\mathbf{a}}(\mathbf{z})\right)\right]^T \tag{44}$$

is the Jacobian of $\mathbf{h}(\mathbf{a})$ evaluated at $\tilde{\mathbf{a}}(\mathbf{z})$. It should also be mentioned that in many cases of interest, such as when the measurement noise is Gaussian, the value of $\mathbf{a}$ that maximises the likelihood is

$$\tilde{\mathbf{a}}(\mathbf{z}) = \mathbf{h}^{-1}(\mathbf{z}) \tag{45}$$

The variance of a uniform pdf in region $I_{\mathbf{a}}(\mathbf{z})$, as an approximation of the variance of $p_1(\mathbf{a};\mathbf{z})$, can be shown to be:

$$\boldsymbol{\Sigma}_{a,1} = \tilde{\mathbf{H}}^{-1}\mathbf{R}\left(\tilde{\mathbf{H}}^{-1}\right)^T \tag{46}$$

Then, for the reasons mentioned above, we propose calculating α_{sp} as:

$$\alpha_{sp} = \frac{\alpha_{max}\gamma \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0}\right)}{\gamma \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0}\right) + (1-\gamma)\,\mathrm{tr}\left(\boldsymbol{\Sigma}_{a,1}\right)} \tag{47}$$

where $\alpha_{max} < 1$ is the maximum value α_{sp} can take, $\gamma \in [0,\, 1]$ is a parameter that controls the weights of the traces

Table I
SINGLE POINT TRUNCATED UNSCENTED KALMAN FILTER UPDATE PHASE STEPS

- Select sigma-points $\mathcal{A}_0^1, ..., \mathcal{A}_0^{N_s}$ (using $\mathbf{x}_{p,0}$ and $\mathbf{P}_{p,0}$).
- Calculate $\tilde{\mathbf{a}}(\mathbf{z})$ and $\mathbf{\Sigma}_{a,1}$ using (46) and (45).
- Calculate α_{sp} using (47).
- Approximate $p_2(\mathbf{a};\mathbf{z})$ by the sigma-points $\left(\mathcal{A}_2^1, ..., \mathcal{A}_2^{N_s+1}\right) = \left(\mathcal{A}_0^1, ..., \mathcal{A}_0^{N_s}, \tilde{\mathbf{a}}(\mathbf{z})\right)$ and weights $\left(w_2^1, ..., w_2^{N_s+1}\right) = \left((1-\alpha_{sp})\, w^1, ..., (1-\alpha_{sp})\, w^{N_s}, \alpha_{sp}\right)$.
- Approximate $\boldsymbol{\mu}_{a,2}$ and $\mathbf{\Sigma}_{a,2}$ using $\left(\mathcal{A}_2^1, ..., \mathcal{A}_2^{N_s+1}\right)$ and $\left(w_2^1, ..., w_2^{N_s+1}\right)$.
- Approximate $\mathbf{x}_{p,2}$ and $\mathbf{P}_{p,2}$ using $\boldsymbol{\mu}_{a,2}$ and $\mathbf{\Sigma}_{a,2}$ in (33), (34) and (35).
- Compute the transformed sigma-points $\mathcal{Z}_2^1, ..., \mathcal{Z}_2^{N_s+1}$ using (39).
- Approximate $\hat{\mathbf{z}}_2$, $\mathbf{S}_2$ and $\mathbf{\Psi}_2$ using (40), (41) and (42).
- Calculate $\hat{\mathbf{x}}_{u,2}$ and $\mathbf{P}_{u,2}$ using $\hat{\mathbf{z}}_2$, $\mathbf{S}_2$ and $\mathbf{\Psi}_2$ in (4) and (8).

of the covariance matrices to select α and $\mathrm{tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. When the information about $\mathbf{a}$ measured from the data is high compared to the prior [11], which means that $\mathrm{tr}(\mathbf{\Sigma}_{a,1})$ is low compared to $\mathrm{tr}(\mathbf{\Sigma}_{a,0})$, $\alpha_{sp} \to \alpha_{max}$. On the contrary, when the information about $\mathbf{a}$ measured from the data is low, which means that $\mathrm{tr}(\mathbf{\Sigma}_{a,1})$ is high compared to $\mathrm{tr}(\mathbf{\Sigma}_{a,0})$, $\alpha_{sp} \to 0$, and the SP-TUKF boils down to a UKF. In general, we have found that using $\gamma = 0.1$ and $\alpha_{max} = 0.8$ attains a high performance although a general detailed analysis regarding the optimal selection of these parameters is left for a future work.

V. SIMULATIONS

In this section, we analyse two aspects of our algorithm. Our paper mainly deals with the update phase of a Kalman filter, therefore, our first simulation is a thorough analysis of the update phase in a 1-D example. We use a 1-D example so that we can easily plot the posterior mean and variances approximated by KF-type algorithms and the Bayes' rule. In addition, we calculate the Kullback-Leibler (KL) divergence, assuming the statistics are Gaussian, between the posterior approximated by the Bayes' rule and the different algorithms. The second simulation is a more realistic one and it is the application of our method to a range-bearing tracking scenario with one target.

A. Analysis of the update phase

The model analysed here represents an amplitude measurement based on the position x of a target assuming the sensor is located on the origin of the coordinate system. The objective is to estimate this position based on the measurement and the prior distribution, which is Gaussian with mean $\hat{x}_{p,0}$ and variance $P_{p,0}$. The measurement noise is zero-mean Gaussian with variance R. The nonlinear function of the measurement model in (1) is

$$h(x) = \begin{cases} A_0 & \text{if} \quad 0 < x \leq d_0 \\ A_0 d_0^n / x^n & \text{if} \quad x > d_0 \\ 0 & \text{elsewhere} \end{cases} \tag{48}$$

where A_0 is the saturation amplitude, d_0 is the saturation distance and n is the signal decay exponent. Note that, according to (48), negative values of x produce a signal of zero amplitude. This would represent a sensor that is only capable of measuring the amplitude of a signal produced by a target coming from a certain direction. Note that the amplitude measurement is presented in one dimension as the measurement function $h(\cdot)$ needs to be bijective at least in a region of the state space, $x > d_0$. In this case, the state is a scalar so the state is not partitioned as in Sections III and IV simplifying the calculations.

Table II
PARAMETERS OF THE 1-D SIMULATION

Parameter	Value
α_{max}	0.8
γ	0.1
n	1
A_0	100
d_0	0.1
R	1
$\hat{x}_{p,0}$	0
$P_{p,0}$	36

Given a measurement $z > 0$, the sigma-point $\tilde{x}(z)$ that represents $p_1(\cdot)$ is:

$$\tilde{x}(z) = \sqrt[n]{\frac{A_0}{z}} d_0 \tag{49}$$

If $z \leq 0$ then $\alpha_{sp} = 0$ and we apply the UKF. Then, assuming that $\tilde{x}(z) > d_0$ and using (47), the parameter α_{sp} is:

$$\alpha_{sp} = \alpha_{max} \frac{\gamma A_0^2 d_0^{2n} n^2 P_p}{\gamma A_0^2 d_0^{2n} n^2 P_p + (1-\gamma) R \tilde{x}(z)^{2n+2}} \tag{50}$$

For different values of the true value of x, we estimate its posterior mean and variance and the KL divergence with respect to Bayes' rule for the UKF, the CKF, the SP-TUKF and the Monte Carlo KF (MCKF) using Monte Carlo simulation with 500 runs. In each Monte Carlo run, the measurement is generated randomly using the measurement equation for the true value of x. The MCKF refers to a KF in which the required moments, given by (5), (6) and (7), are calculated via Monte Carlo sampling from the prior $p_0(\cdot)$. Bayes' rule and the Monte Carlo integration for the MCKF are estimated with 10^4 samples from $p_0(\cdot)$. The parameters used in the simulation are those shown in Table II where we should note that the measurement noise variance is small compared to the prior variance.

The KL divergence with respect to Bayes' rule for the algorithms is shown in Fig. 2. The SP-TUKF approximation is much closer to the real posterior pdf than the MCKF, the UKF and the CKF. Besides, the MCKF and UKF have approximately the same performance and the CKF does not perform properly in this scenario. In addition to the KL divergence, it is also useful to look at the posterior mean and variance, shown in Figs. 3 and 4. When the value of x is small, the SP-TUKF posterior mean is much closer to the real posterior mean than the means given by the MCKF and the UKF. The expected mean of the CKF in this case is very far from the real one. Besides, for small values of x, the posterior variance given by the UKF and the MCKF are very high and it

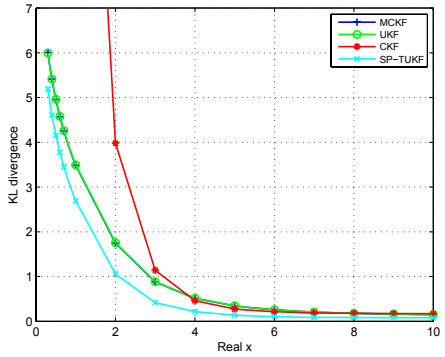

Figure 2. Kullback-Leibler divergence plotted against the real value of x: The SP-TUKF provides the closest approximation of the posterior given by the Bayes' rule.

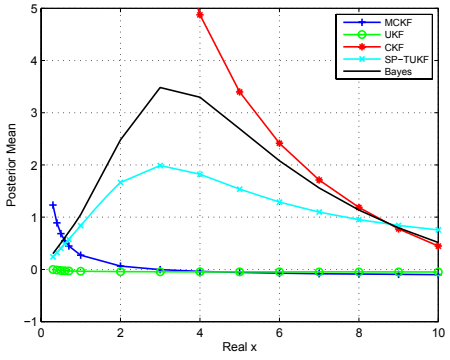

Figure 3. Posterior mean plotted against the real value of x: The SP-TUKF posterior mean represents the real posterior mean more reasonably than conventional KF-type algorithms.

is actually very small. This implies that conventional Kalman-filter-type algorithms do not use the information contained in the measurement properly (when x is small) as the uncertainty of the true posterior is rather small but Kalman's posterior variance is high. When x is high, the algorithms have a roughly similar performance. The SP-TUKF tends to the UKF as the measurement is not very informative and α is low.

Finally, the RMS error averaged over both the parameter and measurement for the algorithms in Table III. As expected, Bayes' rule outperforms the rest of the algorithms followed by SP-TUKF.

Table III
RMS ERROR AVERAGED OVER BOTH THE PARAMETER AND MEASUREMENT

Algorithm	Error
MCKF	5.94
UKF	5.96
CKF	34.35
SP-TUKF	5.67
Bayes' rule	5.22

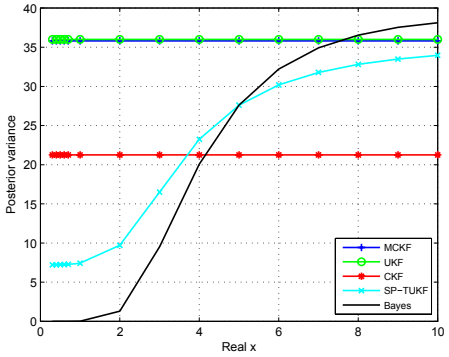

Figure 4. Posterior variance plotted against the real value of x: The SP-TUKF posterior variance represents the real posterior variance more reasonably than conventional KF-type algorithms. Also, the posterior variance of all the conventional KF-type algorithms is independent of the real value of x. This is due to the fact that the approximated posterior variance of KFs is averaged over all the measurements as explained in Section II.

B. Tracking example

In the previous example, we highlighted the improvement in performance of the SP-TUKF over conventional KF-type algorithms only in the update phase. In this example, we are going to show the performance benefits in a dynamic system. We are going to analyse target tracking using range-bearing measurements from a radar located at the origin of the coordinate system. The state vector at time k is $\mathbf{x}^k = \left[\mathbf{a}^k, \mathbf{b}^k\right]^T$ where $\mathbf{a}^k = \left[a_x^k, a_y^k\right]^T$ is the position vector and $\mathbf{b}^k$ is the velocity vector of the target. The measurement model is [2]:

$$z_r^k = \sqrt{\left(a_x^k\right)^2 + \left(a_y^k\right)^2} + \eta_r \tag{51}$$

$$z_\theta^k = \arctan\left(\frac{a_y^k}{a_x^k}\right) + \eta_\theta \tag{52}$$

where $\mathbf{z}^k = \left[z_r^k,\, z_\theta^k\right]^T$, η_r is the measurement noise for the range with variance σ_r^2, η_θ is the measurement noise for the bearing with variance σ_θ^2 and these noises are considered zero-mean Gaussian distributed and independent. We use circular sums and subtractions when performing these operations on the angle measurement, z_θ [12].

In this case, $\tilde{\mathbf{a}}\left(\mathbf{z}^k\right) = \left[z_r^k \cos z_\theta^k,\, z_r^k \sin z_\theta^k\right]^T$. Also, using (46), the approximate trace of the covariance matrix of $p_1\left(\cdot\right)$, which is needed to obtain α_{sp}, is:

$$\mathrm{tr}\left(\mathbf{\Sigma}_{a,1}\right) = \sigma_r^2 + \sigma_\theta^2 \cdot \left(z_r^k\right)^2 \tag{53}$$

The dynamic model of the target is the nearly-constant velocity model [1], [13]:

$$p\left(\mathbf{x}^{k+1} \middle| \mathbf{x}^k\right) = \mathcal{N}\left(\mathbf{x}^{k+1};\, \mathbf{F} \cdot \mathbf{x}^k,\, \mathbf{Q}\right) \tag{54}$$

$$\mathbf{F} = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \otimes \mathbf{I}_2 \tag{55}$$

$$\mathbf{Q} = \sigma_u^2 \begin{pmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{pmatrix} \otimes \mathbf{I}_2 \tag{56}$$

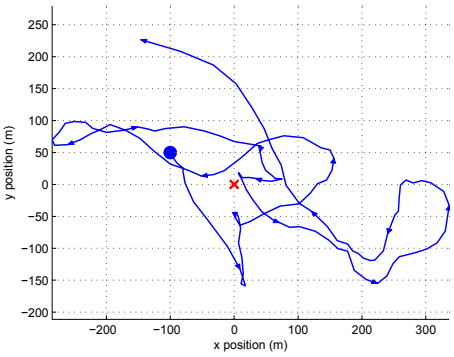

Figure 5. Target trajectory: The initial target position is represented by a blue circle. The target position and direction of movement every 10 times steps are represented by arrows. The radar location is represented by a red cross.

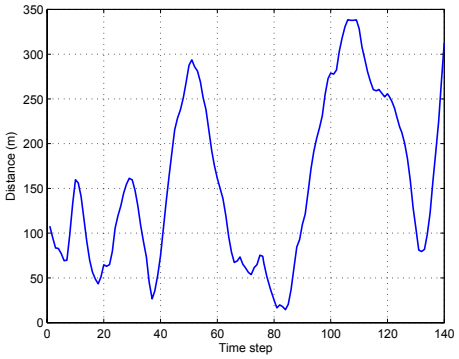

Figure 6. Distance from the target to the radar at each time step.

where $\mathcal{N}(\mathbf{x};\, \overline{\mathbf{x}},\, \mathbf{Q})$ is the Gaussian pdf evaluated at $\mathbf{x}$ with mean $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{Q}$, $\otimes$ is the Kronecker product, τ is the sampling period and σ_u^2 is the continuous-time process noise intensity [1]. One should note that as the dynamics are linear and Gaussian, we can compute the prior at time $k+1$ exactly given a Gaussian approximation at time k. If this were not the case, we could use a UKF prediction step.

The scenario we use to evaluate the performance of the algorithms is represented in Fig. 5. We also show the distance from the radar to the target at each time step in Fig. 6. This is important because conventional KF-type algorithms are expected to perform worse than the SP-TUKF when the target gets close to the radar. This is because the trace of the covariance matrix of $p_1(\cdot)$, given by (53), is low when the target is close to the radar as z_r^k is low.

The sampling period of the trajectory is $\tau = 1\,\mathrm{s}$, $\sigma_u = 10\,\mathrm{m/s}^{3/2}$ and there are $l = 140$ time steps in the simulation. This trajectory corresponds to one realisation of the dynamic system described by (54). We evaluate the tracking performance calculating the RMS of the position error by a Monte Carlo simulation with $m = 200$ realisations.

The prior pdf of the target's state is:

$$\mathbf{x}^0 \backsim \mathcal{N}\left(\mathbf{x}^0;\, \overline{\mathbf{x}}^0, \mathbf{Q}\right) \tag{57}$$

Table IV
PARAMETERS OF THE RANGE-BEARING EXAMPLE

Parameter	Value
α_{max}	0.8
γ	0.1
σ_u	$10\,\mathrm{m/s}^{3/2}$
τ	1 s
σ_r	1 m
σ_θ	$2\pi/180$ rad

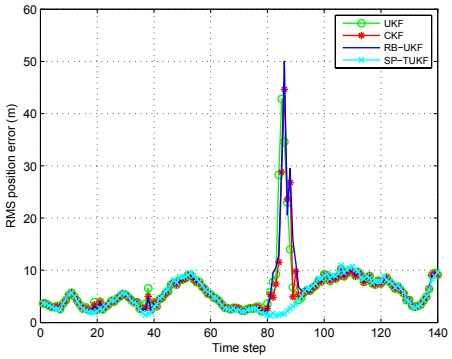

Figure 7. RMS position error plotted against time. When the target is close to the radar, around time steps 40 and 80 see Figure 6, SP-TUKF vastly outperforms conventional KF approximations

where, in each Monte Carlo run, $\overline{\mathbf{x}}^0$ is drawn from a Gaussian distribution whose mean is the real position of the target at time 0 and whose covariance matrix is $\mathbf{Q}$.

We are going to compare the proposed algorithm, SP-TUKF, with three practical algorithms for tracking, the UKF, the CKF and Rao-Blackwellised UKF (RB-UKF) [10]. The parameters of the simulation are shown in Table IV. The RMS position plotted against time is shown in Fig. 7. As expected, UKF, CKF and RB-UKF perform poorly when the target gets close to the radar at around time steps 40 and 80, see Fig. 6. This is due to the fact that the information we obtain from the measurement is very high and conventional KF-type algorithms do not perform properly [6]. In a dynamic system, this can lead to filter divergence as happens at around time 80. In this case, conventional KF-type algorithms manage to recover the track around 10 time steps later. However, SP-TUKF provides a much more accurate estimate of the trajectory as it can deal with situations when the information given by the measurement is very high. This is achieved by adaptively calculating the parameter α_{sp} that allows us to identify when the measurement carries a lot of information and conventional KF-type algorithms fail. Lastly, we show the averaged RMS position error for the algorithms in Table V. For the reasons mentioned above, SP-TUKF is the algorithm that achieves the lowest error.

VI. Conclusions

We have proposed a new method to carry out the update phase of a filtering problem in systems with a nonlinear measurement function, the SP-TUKF. It has low computational burden compared to other methods such as particle filters and

Table V
AVERAGED RMS POSITION ERROR OVER TIME

Algorithm	Averaged RMS position error (m)
SP-TUKF	5.85
UKF	8.19
CKF	8.02
RB-UKF	8.48

can provide a vast improvement in the estimation performance compared to conventional Kalman-filter-type algorithms. This is achieved by approximating the measurement noise pdf by a truncated one and using a modified prior pdf.

Practically, we add one carefully selected sigma-point to the usual set of sigma-points employed by the UKF. The drawback of the current form of the algorithm is that it is not such a general tool as conventional Kalman-filter-type algorithms as its current form requires $\mathbf{h}(\cdot)$ to be a bijective function of some elements of the state.

Once we have put forward the ideas of this way of tackling non-linear problems, the interesting topics of future research are manifold. For instance, one could relax the hypothesis that the measurement function needs to be bijective. In that case the support of $p_1(\cdot)$ could be represented by several disconnected regions. Then, we could add one extra sigma-point per region and represent the posterior by a Gaussian mixture. How to select the weights of the mixture is well-worth exploring.

It is also interesting to represent $p_1(\cdot)$ with more sigma-points. This will slow down the algorithm but better results should be achieved.

VII. ACKNOWLEDGEMENTS

Ángel F. García-Fernández is supported by an FPU Fellowship from Spanish MEC. This work was supported in part by the Spanish National Research and Development Program under Projects TEC2008-02148 and Comonsens (Consolider-Ingenio 2010, CSD2008-00010).

APPENDIX A

In this Appendix, we show that the Bayes' rule applied to $p_2(\mathbf{x};\mathbf{z})$ is the same as applied to $p_0(\mathbf{x})$ in the conditions indicated in Section III. The Bayes' rule applied to $p_2(\mathbf{x};\mathbf{z})$, denoted as $l(\mathbf{x}|\mathbf{z})$, is

$$l(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})\, p_2(\mathbf{x};\mathbf{z}) \tag{58}$$

Substituting (12), (18) and (19) into (58), we get

$$\begin{aligned} l(\mathbf{x}|\mathbf{z}) \propto & p_{\boldsymbol{\eta}}(\mathbf{z}-\mathbf{h}(\mathbf{a}))\, \chi_{I_{\mathbf{x}(\mathbf{z})}}(\mathbf{x}) \cdot \\ & \left[\alpha_{sp}\frac{1}{\varepsilon_1} p_0(\mathbf{x})\, \chi_{I_{\mathbf{x}(\mathbf{z})}}(\mathbf{x}) + (1-\alpha_{sp})\, p_0(\mathbf{x})\right] \end{aligned} \tag{59}$$

$$= p_{\boldsymbol{\eta}}(\mathbf{z}-\mathbf{h}(\mathbf{a}))\, \chi_{I_{\mathbf{x}(\mathbf{z})}}(\mathbf{x})\, p_0(\mathbf{x}) \left(\alpha_{sp}\frac{1}{\varepsilon_1} + 1 - \alpha_{sp}\right) \tag{60}$$

Using (12) in (60)

$$\begin{aligned} l(\mathbf{x}|\mathbf{z}) &\propto p(\mathbf{z}|\mathbf{x})\, p_0(\mathbf{x}) \left(\alpha_{sp}\frac{1}{\varepsilon_1} + 1 - \alpha_{sp}\right) \\ &\propto p(\mathbf{z}|\mathbf{x})\, p_0(\mathbf{x}) \end{aligned} \tag{61}$$

Therefore, (61) proves that the Bayes' rule remains unaltered if $p_2(\mathbf{x};\mathbf{z})$ is the prior instead of $p_0(\mathbf{x})$.

APPENDIX B

In this Appendix, we calculate the mean of $p_2(\mathbf{a},\mathbf{b};\mathbf{z})$ considering we know the first two moments of $p_0(\mathbf{a},\mathbf{b})$, given by (20) and (21), and the first two moments of $p_2(\mathbf{a};\mathbf{z})$, given by (31) and (32). Firstly, we must note that the truncation only affects the part of the pdf that corresponds with $\mathbf{a}$, see (14). Assuming that $p_0(\mathbf{a},\mathbf{b})$ is Gaussian [1]

$$\boldsymbol{v}_b(\mathbf{a}) \triangleq \mathsf{E}[\mathbf{b}|\mathbf{a}] = \boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1}(\mathbf{a}-\boldsymbol{\mu}_{a,0}) \tag{62}$$

Using (24), the mean $\boldsymbol{\mu}_{b,2}$ is

$$\begin{aligned} \boldsymbol{\mu}_{b,2} &= \int \mathbf{b} p_2(\mathbf{a},\mathbf{b};\mathbf{z})\, \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} = \int \mathbf{b} p_0(\mathbf{b}|\mathbf{a})\, p_2(\mathbf{a};\mathbf{z})\, \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \\ &= \int \mathsf{E}[\mathbf{b}|\mathbf{a}]\, p_2(\mathbf{a};\mathbf{z})\, \mathrm{d}\mathbf{a} \end{aligned} \tag{63}$$

Using (62) and (31) in (63):

$$\boldsymbol{\mu}_{b,2} = \boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1}(\boldsymbol{\mu}_{a,2} - \boldsymbol{\mu}_{a,0}) \tag{64}$$

REFERENCES

[1] Y. Bar-Shalom, T. Kirubarajan, and X. R. Li, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2002.
[2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
[3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.
[4] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
[5] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.
[6] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
[7] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.
[8] M. Briers, "Improved Monte Carlo methods for state-space models," Ph.D. dissertation, University of Cambridge, 2007.
[9] I. Arasaratnam, S. Haykin, and T. Hurd, "Cubature Kalman filtering for continuous-discrete systems: Theory and simulations," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 4977–4993, Oct. 2010.
[10] M. R. Morelande and B. Moran, "An unscented transformation for conditionally linear models," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, April 2007, pp. 1417–1420.
[11] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.
[12] N. Nikolaidis and I. Pitas, "Nonlinear processing and analysis of angular signals," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3181–3194, Dec. 1998.
[13] A. F. García-Fernández and J. Grajal, "Multitarget tracking using the joint multitrack probability density," in *12th International Conference on Information Fusion*, July 2009, pp. 595–602.